Computational Portals and the Berkeley Water Center

Deb Agarwal (UCB/LBNL)

Science Advisory Board Mtg, LBNL, April 24, 2006

Distributed Data Sets

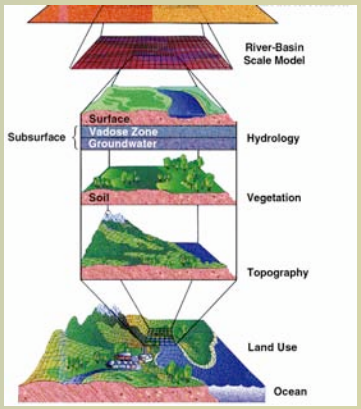
USGS science for a changing world

CIMIS

NAS

DEPARTMENT OF WATER RESOURCES

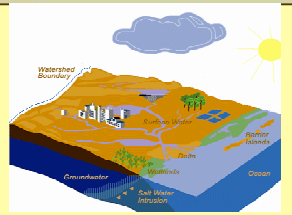
Data Harvesting and Transformations

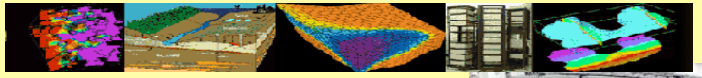


BWC Data Gateway

BWC Analysis Gateway

Data Cleaning, Models, Analysis Tools





Computational Resources

BWC Water Portal

Knowledge discovery, Hypothesis testing, Water Synthesis

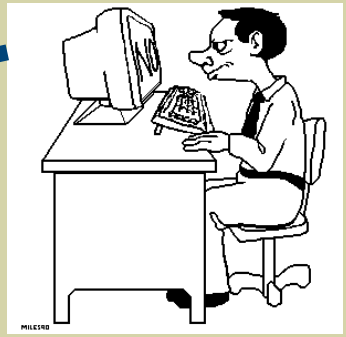
Dissemination and Archiving

Microsoft

Building BWC Water Cyberinfrastructure to Connect Data, Resources, and People

Microsoft Project – BWC Computational Portal

- Single point of access to donated computational resources including: NERSC, SDSC, and Millenium cluster
- Provide grid authentication to resources without requiring user certificates
- Provide access to BWC data resources



Distributed Data Sets

USGS science for a changing world

CIMIS

NAS

DEPARTMENT OF WATER RESOURCES

DOTNETNUKE™
community • content • collaboration

CredEx

Data Cleaning, Models, Analysis Tools

Effective Use of Computing Requires

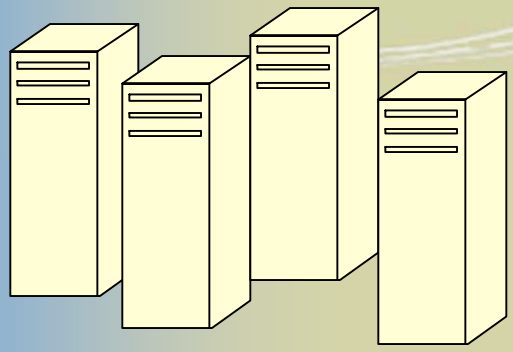
- Shared accounts
  - Portal authenticates individual users
  - Portal uses a single account on resources with audit tracking to allow accounting and cybersecurity mapped to individuals
- Resource decisions
  - Location of data
  - Location of models and analysis routines
  - Compatability of resources

Deliverables

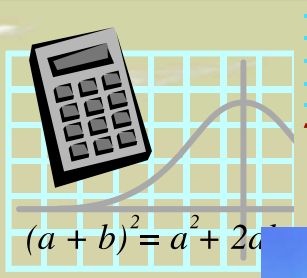
- Computational portal design (4 months)
  - Access to data
  - Ability to specify analysis and modeling jobs
  - User authentication using username/password
- Prototype shared account capability on compute resources (3 months)
- Ability to choose resource to use (1 month)
- Optimization suggestions provided (4 months)

Carbon Climate Community Workbench Proposal to NSF

- Build a problem solving environment oriented to the scientist
- Easy access to a wide range of data sources, analysis routines, and models
- High-level interface allowing visual composition of complex workflows
- Knowledge-level representation of process
- Visual presentation of data and results
- Web services-based infrastructure



Data Providers:
Host Ameriflux
Climate Data
Statsgo Soils Data
MODIS products



Tools:
Statistical
Graphical

Microsoft

Web Service Interface to Data and Tools

Web-based Workbench access

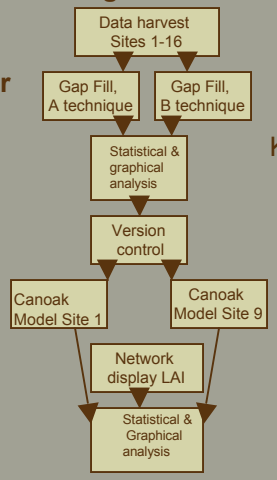
Choose Ameriflux Area/Transect, Time Range, Data Type

LAI
Temp
Fpar
Veg Index
Surf Refl
NPP
Albedo

Import other Datasets

Climate
Statsgo
MODIS

Design Workflow



Ecology Toolbox

Data Cleaning Tools

Knowledge Generation Tools

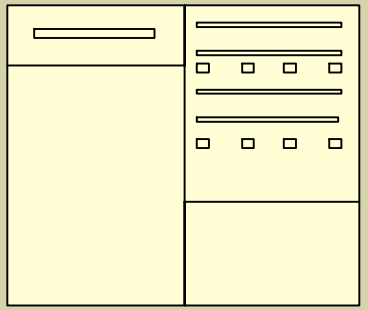
Data Mining and Analysis Tools

Modeling Tools

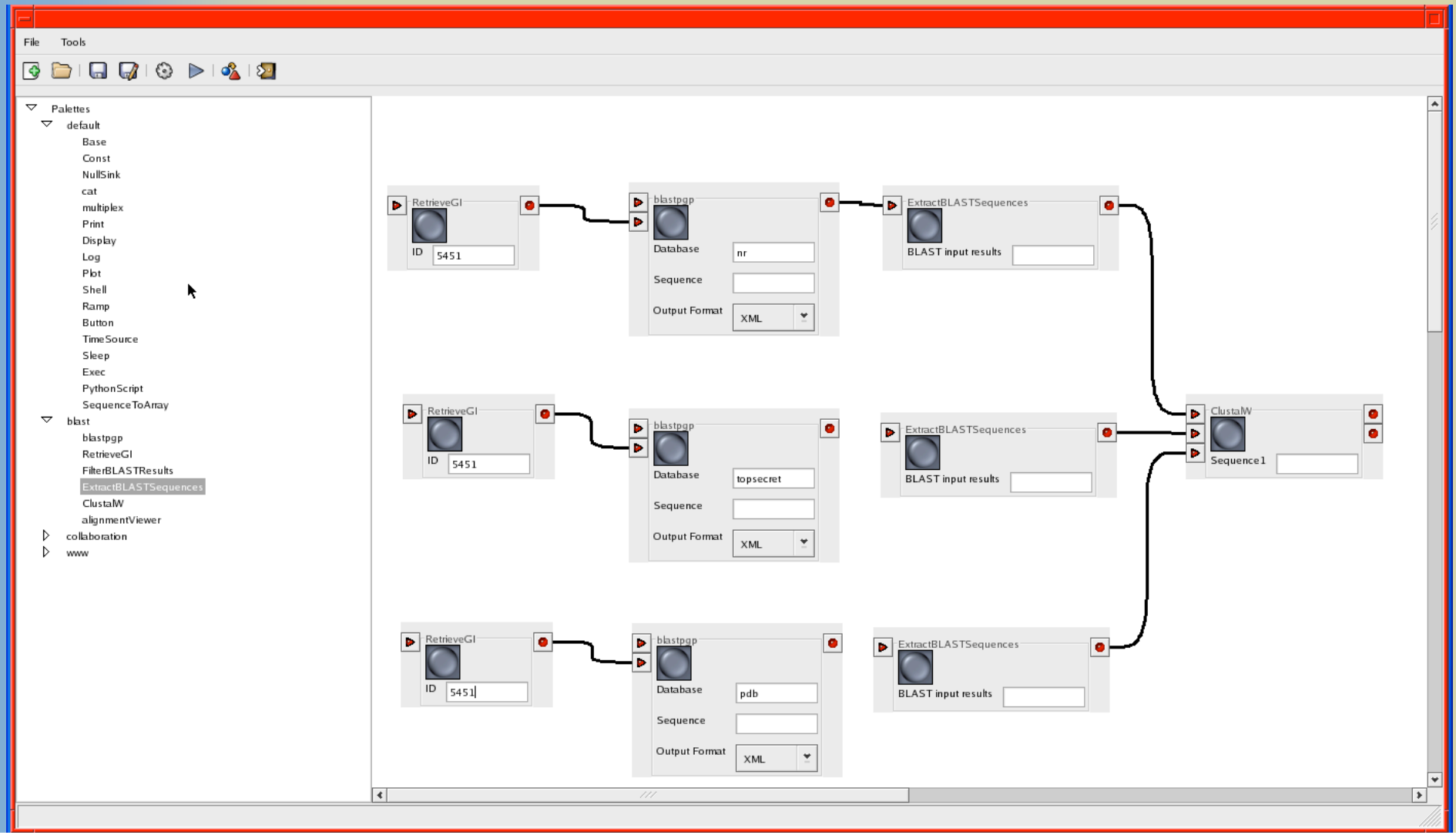
Visualization Tools

Carbon Community Workbench

Compute Resources



ViCE Workflow Interface



Berkeley Water Center Infrastructure

- Data
  - Uniform and easy access to a wide variety of sources
  - Ability to clean and analyze data
- Compute resources
  - Analysis routines
  - Models
- Prototype high-level problem solving environment to support synthesis

Current Status of infrastructure

- Campus server operational with
  - Windows 2003
  - SQL Server – with Ameriflux data loaded
  - Proclarity
  - IIS 6.0
  - Visual Studio .NET
  - DotNetNuke
- DotNetNuke configured to access local SQL Server Ameriflux data
- Identical server installed at LBL for prototyping and testing